Il progetto National Bibliography Number Italia (NBN:IT). Un identificatore persistente a supporto del deposito legale nazionale delle risorse digitali

Emauele Bellini, Chiara Cirinnà, Maurizio Lunghi, Cinzia Luddi, Maurizo Messina, Giovanni Bergamin, Raffaele Messuti, Giovanna Cordani, Roberto Delle Donne, Francesca Rossi, Marialaura Vignocchi, Stefania Arabito, Jordan Piščanc'

Introduzione

L'associazione di un Persistent Identifiers (PI) a una risorsa digitale è considerata oggi una best practice che contribuisce alla risoluzione del problema del reperimento affidabile delle risorse in rete, andando ben oltre la soluzione dell'errore Internet 404. Attualmente i PI sono considerati un mezzo per poter garantire non solo l'identificazione e l'accesso nel lungo periodo, ma anche la certificazione, l'autorevolezza e l'affidabilità delle risorse stesse a patto però che l'implementazione dell'infrastruttura prenda in considerazione non

JLIS.it. Vol.3, n.1 (Giugno/June 2012).
DOI: 10.4403/jlis.it-4789

solo i requisiti tecnici, ma soprattutto quelli organizzativi, di processo, di comunità, di conservazione etc. L'associazione di un PI a una risorsa digitale può essere usata come un concreto strumento di localizzazione affidabile, ma, allo stesso tempo, può supportare la certificazione dell'autenticità della risorsa, della sua provenienza, dei suoi diritti di sfruttamento etc.

Ovviamente la garanzia di una reale persistenza degli identificativi è principalmente legata all'impegno mostrato da un'organizzazione nell'assegnarli e gestirli. Questo impegno deve declinarsi nell'adozione di precise politiche di servizio, nell'implementazione di specifiche soluzioni tecnologiche, nella definizione di un modello di business sostenibile, nell'identificazione della comunità di riferimento, nel rispetto di opportuni standard condivisi internazionalmente. La reale persistenza di un sistema di identificazione si deve quindi basare sul rispetto di una serie di criteri qualitativi e quantitativi verificabili così come proposto dai risultati di iniziative internazionali, quali NESTOR[1] («NESTOR Catalogue of criteria for assessing the trustworthiness of PI systems») e APARSEN[2] (Bazzanella et al.).

Attualmente sono disponibili diverse tecnologie e standard per l'implementazione di un sistema di attribuzione di PI; manca tuttavia un accordo sulla loro adozione, anche perché alcuni di questi sistemi sono nati come soluzioni tecniche senza il supporto delle comunità di utenti che si aspettano dai PI livelli di servizio. Siste-

[1]NESTOR, http://www.langzeitarchivierung.de/eng, è un network di competenza tedesco sulla digital preservation e ne fanno parte biblioteche, archivi, musei e esperti del settore.

[2]Alliance for Permanent Access to the Records of Science Network, http://www.alliancepermanentaccess.org/index.php/current-projects/aparsen, è un network of excellence sulla digital preservation finanziata dalla commissione europea che si occupa di studiare tuti gli aspetti della conservazione dei record scientifici, dagli standard alla loro identificazione, dalle soluzioni di storage alla creazione di un centro di competenza virtuale europeo.

mi come il PURL[3] oppure i Cool URI[4] (Berners-Lee) presentano indubbiamente notevoli vantaggi (il più rilevante tra questi è la loro immediata dereferenziabilità[5] attraverso il protocollo HTTP[6]), ma il limite di queste soluzioni è che la loro persistenza non è in principio garantita da una terza parte indipendente e autorevole. Inoltre, nell'ambito del progetto europeo APARSEN è stata condotta un'indagine i cui risultati mostrano che un numero limitato di approcci, come il Digital Object Identifier DOI (33%), il sistema Handle (29%) e gli identificatori URN-based (Moats) (25%), sembrano ormai prevalere a livello europeo (Bazzanella et al.).

Una particolare attenzione merita in ogni caso il Cool URI. Com'è noto il Cool URI è un approccio alla persistenza delle URL basato sulla loro progettazione. Questo approccio principalmente tecnico è considerato una best practice dal W3C per l'implementazione delle URI del semantic web e dei Linked Data.[7] L'assunzione di base è che una progettazione corretta delle URI dovrebbe ridurre i casi in cui risulterà necessario cambiarle, garantendone la loro stabilità nel tempo. Un esempio di questa best practice è quello di evitare di esplicitare l'estensione delle pagine web come .php o .asp, in modo che se cambia la tecnologia di base (da PHP ad ASP) la URI rimane

[3]OCLC Persistent URL è un servizio che consente agli utenti di generare delle URL con certe caratteristiche di persistenza riferendo a queste le URL delle proprie risorse considerate più soggette a cambiamenti, migrazioni, ecc. URL http://purl.oclc.org.

[4]Cool URI è una best-practice del W3C che intende rendere persistenti le URL tramite l'adozione di una serie di regole e pratiche di gestione in grado di ridurre i casi in cui risulterebbe necessario cambiarle o dismetterle, http://www.w3.org/Provider/Style/URI.html.

[5]La locuzione "dereferenziare una URI" è riferita alla capacità di recuperare la rappresentazione della risorsa identificata da quella URI tramite una applicazione (es. browser). In altre parole è la capacità di accedere ad esempio a una pagina web, o a un PDF all'intero di un browser tramite una URL.

[6]http://www.ietf.org/rfc/rfc2616.txt.

[7]http://www.w3.org/TR/2007/WD-cooluris-20071217, http://www.w3.org/DesignIssues/LinkedData.html.

la stessa. In questa ottica, la persistenza si baserebbe esclusivamente sull'impegno della singola istituzione; s'instaura così una relazione di trust diretta tra l'utente e l'istituzione stessa, senza mediazioni. In questo senso il trust è il livello di fiducia che un utente ha nella capacità di un'istituzione di mantenere le caratteristiche dichiarate del servizio offerto. Il livello di fiducia viene ciclicamente verificato durante l'interazione che l'utente ha con il servizio stesso, e dove si ha una conferma positiva, il trust si rafforza; in caso contrario decresce. È ben noto, purtroppo, che l'impegno della singola istituzione non è più sufficiente a garantire né la persistenza della URL né tanto meno la certificazione delle risorse in termini di provenienza, autenticità, integrità, conservazione etc. Nella pratica, le risorse si muovono sulla rete, vengono modificate o cancellate a causa di una infinità di fattori che non possono essere sempre predeterminati dalle politiche di gestione dei contenuti delle singole istituzioni, né tanto meno governate da best practice esclusivamente tecniche. Un caso tipico è quando un' istituzione chiude le sue attività perché viene comprata da un'altra istituzione o viene soppressa, oppure semplicemente cambia nome. In questi casi gli oggetti possono essere rinominati per essere adattati al workflow interno della nuova istituzione, o ceduti ad altre istituzioni, o nel peggiore dei casi cancellati perché non più corrispondenti agli obiettivi istituzionali. È evidente che tutte queste azioni possono produrre l'invalidazione delle vecchie URL indipendentemente da come queste siano state costruite. Questo potrebbe non essere un problema se l'istituzione non gestisce contenuti di valore scientifico, culturale o amministrativo, ma cosa diversa è se la medesima sorte la subiscono istituzioni come editori scientifici, biblioteche, archivi, etc. In questi casi, ad esempio, bibliografie basate su semplici URL o anche Cool URI, facenti riferimento a risorse che erano presenti negli archivi di queste istituzioni, non potranno più essere usate, ad esempio, per verificare il lavoro scien-

tifico o per calcolare indici bibliometrici. La cosa più critica, però, è l'impossibilità di implementare sistemi di verifica dell'autenticità, della provenienza e dell'integrità di queste risorse in assenza di una terza parte che si faccia garante dell'associazione nome – risorsa.

L'uso dei PI consente, quindi, di inserire non solo nel web ma anche nel nuovo web semantico quel necessario livello di trust che attualmente manca allo scenario così come auspicato dai risultati del seminario sui PI e Linked Open Data confluiti nel Den Haag Manifesto,[8] a cui la Fondazione Rinascimento Digitale ha contribuito attivamente. Questo livello di trust è implementato attraverso l'identificazione di una terza parte indipendente e autorevole (come ad esempio le Biblioteche Nazionali) con un preciso mandato che tenga in considerazione non solo i requisiti tecnici, ma anche quelli di comunità (ad esempio la granularità dell'identificativo), organizzativi (autorevolezza della Registration Authority riconosciuta nel dominio di applicazione, impegno sulla durata del servizio) e di processo (policy di dominio sulla selezione delle risorse, etc.) (Bellini et al., «Semantics-Aware Resolution of Multi-part Persistent Indentifiers»; «Persistent Identifiers distributed system for Cultural Heritage digital objects»).

Il progetto NBN Italia

Il progetto NBN Italia (NBN:IT) ha inteso realizzare un servizio di identificazione persistente basato sullo standard NBN che assegna alle biblioteche nazionali la responsabilità esclusiva della gestione

[8] http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fPersID%2fWorkshop+POID%2fDen+Haag+Manifesto+20110825_2.pdf. Si noti in ogni caso la differenza tra l'identificatore di tipo URL appena citato e gli identificatori riportati alle precedenti, costruiti come Cool URI.

del sottodominio nazionale NBN:IT. Partendo dall'attuale orientamento per i sistemi NBN, derivato da progetti come PersID, da report come CENL Task Force on Persistent Identifiers e dalle raccomandazioni contenute ne The National Libraries Resolver Discovery Service («CENL Task Force on Persistent Identifiers, Report 2007»), dove se ne sottolinea il ruolo a supporto della digital preservation, in Italia si è consolidata l'idea di legare NBN:IT a un servizio che è istituzionalmente affidato alle biblioteche nazionali: il deposito legale. I primi studi intorno al NBN sono cominciati nel 2007 su iniziativa della Fondazione Rinascimento Digitale con il prezioso supporto dell'Università Statale di Milano con significativi risultati dal punto di vista della ricerca (Bellini et al., «Persistent Identifiers distributed system for Cultural Heritage digital objects»; «Semantics-Aware Resolution of Multi-part Persistent Indentifiers»; Bellini e Ceravolo). Le attività sono poi proseguite con una prima fase di sperimentazione che ha visto lo sviluppo dell'architettura e il test di vari prototipi (Bellini et al., «Persistent Identifier Distributed System for Digital Libraries, Information Technology»; «Persistent identifiers: jNBN, a JEE application for the management of a national NBN infrastructure»). Infine, in considerazione della vocazione di NBN verso la digital preservation, il progetto NBN:IT è confluito nell'ambito del Deposito Legale nazionale, diventandone oggi parte integrante. L'iniziativa NBN vede oggi coinvolti il consorzio di Magazzini Digitali (Vignocchi, Bergamin e Messuti; «Le linee guida per l'accesso aperto alle tesi di dottorato») per il deposito legale composto dalla Biblioteca Nazionale Centrale di Firenze (BNCF), la Biblioteca Nazionale Centrale di Roma (BNCR), la Biblioteca Nazionale Marciana di Venezia e la Fondazione Rinascimento Digitale (FRD) e la CRUI che sovraintende la sperimentazione con le tesi di dottorato. Attualmente il progetto NBN è in fase avanzata di sviluppo e i suoi principali obiettivi possono essere riassunti nei seguenti punti:

1. la creazione di un registro nazionale di nomi stabile e affidabile come strumento a supporto del deposito legale per la certificazione degli oggetti digitali prodotti dalle istituzioni culturali, scientifiche e educative,

2. l'accesso permanente alle risorse digitali prodotte dalle istituzioni culturali italiane, incluso il materiale digitalizzato o non ancora pubblicato,

3. la condivisione di politiche per la conservazione di lungo periodo degli oggetti digitali,

4. la sostenibilità dei costi del servizio e le responsabilità di gestione dei nomi da parte delle istituzioni.

NBN: un identificatore a supporto del deposito legale

La decisione di associare un servizio di identificazione persistente a un servizio di deposito legale ha delle implicazioni importanti, che meritano qualche considerazione. La definizione del nuovo servizio NBN prende avvio dalle considerazioni espresse nel report CENL Task Force on Persistent Identifiers [2] e nelle CENL Recommendation - The National Libraries Resolver Discovery Service (RDS) («CENL Task Force on Persistent Identifiers, Report 2007») e si consolida nell'idea di collegare NBN a un servizio specifico, andando ben oltre la sola esigenza di identificazione e realizzando un sistema a valore aggiunto. Già nel 2007 i report del CENL promuovevano l'adozione di NBN da parte delle Biblioteche Nazionali, scoraggiando l'uso di altri sistemi. Inoltre, l'idea che NBN fungesse anche da risolutore di ultima istanza di DOI, una volta che una risorsa digitale non fosse stata più accessibile dall'archivio dell'editore, ci fa comprendere la differenza e la complementarietà di NBN rispetto agli altri sistemi. L'iniziativa italiana ha recepito in pieno queste

indicazioni spingendo le implicazioni di queste scelte ben oltre la sfera applicativa. Infatti, il conferimento all'identificativo NBN di significati ulteriori, a nostro avviso, consente di comunicare quel valore di "fiducia" atteso dagli utenti dei servizi erogati dalle pubbliche amministrazioni in generale, e, in particolare, dei servizi di conservazione e identificazione.

Il ruolo iconico di NBN

Un identificatore persistente utilizzato da organismi istituzionali per identificare un'entità o un oggetto prodotto dall'organizzazione stessa come risultato di un processo amministrativo e burocratico ha un potenziale ruolo iconico nella definizione dei requisiti di trust nell'information society («Final Report of the Project Team of CEN/ISSS eGovernment Focus Group, CEN»; Winer; Wilson). Luhmann («Trust: a mechanism for the reduction of social complexity») definisce il trust come un meccanismo per ridurre l'incertezza, la complessità e il rischio che percepiscono gli utenti riguardo all'entità in esame. In quest'ottica, quindi, gli utenti si aspettano che l'output di un sistema ritenuto affidabile, sia corretto e in linea con le loro aspettative. La reiterazione di questa conferma contribuisce a stabilizzare questa fiducia. L'attribuzione di un identificativo a una risorsa, assegnato da parte di una pubblica amministrazione, dovrebbe implicare il superamento da parte della risorsa stessa di un processo ufficiale di riconoscimento (carta d'identità), di presa in carico (numero di protocollo), di tracciamento (codice fiscale, targa), e così via. La possibilità di verifica dell'avvenuto processo è garantita non solo dal mantenimento in appositi registri (anagrafe, protocollo, etc.) dell'associazione tra l'identificativo assegnato e l'entità oggetto del processo, ma anche dalla loro immediata accessibilità a fini di controllo. Quindi, la sola presenza di un identificativo di questo tipo conferisce all'entità a cui è assegnato, un livello di trust supe-

riore rispetto a entità della stessa natura che ne sono sprovviste. Il ruolo iconico di un identificatore deriva, quindi, dal fatto che la sua assegnazione (o adozione) è subordinata al superamento di un processo le cui modalità di attuazione sono conosciute, accettate e formalmente e/o legalmente disciplinate. In accordo a questa visione si è ritenuto inevitabile dotare il processo di deposito legale di un sistema di identificazione delle risorse stesse. L'associazione di un NBN a una risorsa certificherà pertanto all'utente che quella stessa risorsa è stata legalmente depositata (o che è in corso il deposito legale) in accordo alle regole che disciplinano questo processo. La dimostrazione dell'avvenuto processo può essere perpetrata solo tramite la risoluzione e confermata dall'utente finale.

Il rapporto tra NBN e DOI/Handle

Dopo un'iniziale proliferazione di tecnologie e proposte di standardizzazione, attualmente il dominio sembra pragmaticamente focalizzato su soluzioni come DOI, Handle e identificativi URN based come NBN. I sistemi tra loro non sono comunque equivalenti. Infatti, sia Handle che DOI (che si basa su l'infrastruttura Handle) possono essere assegnati a qualunque tipo di risorsa e la loro funzione principale è quella di identificazione (per Handle) o di indirizzamento alle pagine degli editori dove è possibile acquistare la risorsa (per DOI). Nessuno dei due però può dirsi strettamente subordinato a un workflow/processo specifico e certificato come avviene ad esempio per l'assegnazione di ISBN (processo di pubblicazione) o per il rilascio della carta di identità (processo di riconoscimento). Infatti, anche se il DOI è generalmente usato dagli editori nella fase finale del processo di pubblicazione, chiunque può acquisire tale identificativo commerciale per le proprie risorse senza che queste ultime siano sottoposte necessariamente a un processo di selezione, catalogazione, archiviazione, etc. Il DOI, come Handle, è quindi un

identificativo "generico" (avere un DOI o un Handle di per sé, non dice niente di più del fatto che la risorsa sia identificata su Internet) e pertanto supporta l'erogazione di altri tipi di servizi come quelli citazionali, che non sono in alternativa a quelli forniti attraverso NBN, ma vi si aggiungono, completando l'offerta. Infine, sia per il DOI che per Handle, la responsabilità della stabilità delle risorse e della loro conservazione è interamente affidata al content provider, quindi, nel caso l'istituzione non abbia adottato specifiche politiche di conservazione, il rischio che nel lungo periodo il DOI non risolva più correttamente è concreto. Diversamente, NBN associato al servizio di deposito legale assicura un'affidabilità maggiore in quanto beneficia dell'assunzione di responsabilità della pubblica amministrazione di conservare nel lungo periodo le risorse. In quest'ottica, la garanzia di risoluzione di un NBN nel lungo termine è data dal fatto che questo punterà sempre almeno alla copia conservata nel deposito legale.

Attualmente le risorse possono già ricevere più di un identificativo durante il loro ciclo di vita. Ad esempio, un articolo pubblicato su una rivista non Open Access e depositato in un repository istituzionale OA basato su un sistema DSpace il cui dominio è registrato al CNRI, avrà come minimo un ID interno OAI (non risolvibile), un Handle (risolvibile alla risorsa depositata nel repository) e un DOI (risolvibile principalmente alla copia depositata presso l'editore), quindi l'introduzione di un nuovo identificativo con le caratteristiche precedentemente descritte non potrà che rappresentare un valore aggiunto per la risorsa stessa. Infatti, questi identificativi, tutti gestiti dai metadati descrittivi associati alla risorsa, di per sé non sono in contrasto in quanto, come abbiamo detto, assolvono a servizi diversi, e il NBN rappresenta un metadato aggiuntivo in grado di informare l'utente che esiste una copia della risorsa legalmente depositata alla biblioteca nazionale. La prova del deposito è

ottenibile tramite il servizio di risoluzione del nome.

Lo standard IETF RFC 3188 – National Bibliography Number

Il PI definito come National Bibliography Number (NBN) (Hakala) nasce nel contesto del ruolo e dalle responsabilità che le biblioteche nazionali hanno nel campo del controllo bibliografico universale (in particolare con le bibliografie nazionali).[9] Si tratta di uno standard basato su Uniform Resource Name (URN) (Masinter e Sollins; Moats; Daigle et al.) e definito dall'RFC 3188 in ambito IETF. Lo standard NBN riprende molti degli aspetti definiti nell'RFC 2288 (Lynch, Preston e Daniel) nel quale si fa riferimento all'utilizzo degli identificatori bibliografici come l'International Standard Book Number (ISBN), l'International Standard Serials Number (ISSN) («NISO/ANSI/ISO 2108:1992 Information and documentation - International standard book number (ISBN)»; «ISO 3297:1986 Documentation - International standard serial numbering (ISSN)»), e il Serial Item and Contribution Identifier (SICI) («NISO/ANSI Z39.56-1997 Serial Item and Contribution Identifier»). I namespace basati sullo standard URN (quindi anche NBN) sono fortemente orientati alla sintassi dei nomi e alle responsabilità della gestione dei sub-namespace. Infatti, esistono diversi gradi di libertà per l'implementazione del servizio, che può essere diverso per comunità diverse. Nel caso di NBN, lo standard prevede una ripartizione di responsabilità simile ai domini DNS dove è previsto un suffisso che identifica il paese di riferimento (es. NBN:IT:xxxx oppure NBN:DE:yyyy).

Quindi la struttura tipica di un NBN è la seguente:

URN:NBN:<country code>:<sub-namespace>-<assigned NBN>

[9] Per il concetto di Controllo Bibliografico Universale si veda Solimine (*Controllo bibliografico universale*).

dove ″country code″ è il codice del Paese, indicato secondo lo standard ISO 3166, il ″sub-namespace″ è il codice del sottodominio e ″assigned NBN″ è la stringa di caratteri scelta.

La responsabilità per la gestione del sottodominio nazionale è delegata, come previsto dallo standard, alle biblioteche nazionali di ogni paese (Hakala). NBN è stato registrato al registro Internet Assigned Numbers Authority (IANA) degli URN nell'ambito del Nordic Metadata Project su proposta della Conference of Directors of National Libraries (CDNL) e Conference of European National Libraries (CENL). Il primo obiettivo dello standard era quello di avere un sistema di identificazione per tutte quelle risorse digitali che non possedevano un identificativo perché fuori da circuiti di produzione specifici, come avviene invece per i libri con ISBN o i periodici con ISSN. Esempi di queste risorse che andavano a collocarsi nella cosiddetta letteratura grigia sono le tesi di dottorato, i rapporti tecnici, gli oggetti digitalizzati come i manoscritti antichi etc. Le attuali iniziative di NBN in Germania[10] come in Olanda, ad esempio, pur mantenendo una certa autonomia nell'implementazione del servizio (da qui il tentativo di avere policy e un sistema di risoluzione comuni a livello europeo sperimentato dal progetto PersID[11]), hanno una chiara vocazione verso la digital preservation delle risorse identificate. Questa relazione è forte a tal punto che il riferimento alla conservazione delle risorse digitali identificate dal namespace NBN sarà esplicitato nella prossima versione dell'RFC 3188. Nel solco di questa visione si colloca l'attività della Fondazione Rinascimento Digitale nell'ambito del gruppo di lavoro IETF[12]

[10] http://www.persistent-identifier.de/english/3352-EPICUR_policy.php.

[11] PersID- Building a persistent identifier infrastructure, http://www.persid.org. Il progetto PersID, finanziato dalla SURF Foundation, aveva come scopo la progettazione di un metaresovler tra i vari sistemi NBN europei.

[12] L'IETF ha attivato un gruppo di lavoro aperto per la revisione dello standard 3188.

per la revisione dell'RFC 3188 e di gruppi di progetto internazionali come PersID e la collaborazione con il Ministero dei beni Culturali (MiBAC) e la Conferenza dei Rettori delle Università Italiane (CRUI) per la realizzazione di un sistema analogo nel nostro paese.

Il servizio NBN Italia

Sulla base dei risultati raggiunti dalle prime ricerche sull'argomento, l'attuale gruppo di lavoro ha definito a livello generale l'organizzazione del sistema e le principali funzionalità. Ovviamente, l'iniziativa potrà prevedere evoluzioni e ottimizzazioni sulla base dei requisiti della comunità, che possono variare a partire dai livelli di performance del sistema fino alla distribuzione delle responsabilità dell'assegnazione dei nomi, come già ipotizzato (Bellini et al., «Persistent Identifiers distributed system for Cultural Heritage digital objects»; «Semantics-Aware Resolution of Multi-part Persistent Indentifiers») e in parte sperimentato («Persistent Identifier Distributed System for Digital Libraries, Information Technology»; «Persistent identifiers: jNBN, a JEE application for the management of a national NBN infrastructure»). In questa fase, considerato il contesto di start up, si è privilegiato un'organizzazione più centralizzata per esigenze di controllo e assessment del sistema.

Il livello di persistenza

NBN si propone un livello di persistenza molto alto. Nel contesto URN si possono identificare almeno tre livelli di persistenza:

1. persistenza dell'identificatore. Se la risorsa non è più disponibile in rete, l'identificatore URN può continuare ad avere senso (ad esempio come prova che in un determinato momento quella risorsa esisteva);

2. persistenza dell'abbinamento URN e URL. Si tratta di un impegno che garantisce nel lungo periodo che un URN sia sempre risolto (porti almeno a un indirizzo di tipo URL). Non viene garantita l'accessibilità della risorsa ma viene garantito – in caso di non disponibilità permanente della risorsa in rete – l'accesso alla c.d. tombstone (es. "questo ebook non è più in commercio");

3. persistenza della Risorsa referenziata da NBN. Assicurare nel lungo periodo l'esistenza e accessibilità della risorsa referenziata da URN. Questo è il livello di persistenza di NBN reso possibile solo dal deposito (legale o volontario) presso le biblioteche nazionali e dalla descrizione autorevole della Bibliografia Nazionale.

L'organizzazione

Il servizio NBN propone un'organizzazione su due livelli:

- il Registro Centrale dei nomi (RC) gestito dalla BNCF
- i Content Provider (CP), cioè coloro che hanno le risorse che devono essere inviate verso il deposito legale.

La richiesta di assegnazione del NBN avviene in modo decentrato direttamente dai CP autorizzati, seguendo un certo criterio di selezione delle risorse e di accreditamento dei CP stessi. Le istituzioni accreditate riceveranno un sub-namespace (in genere l'acronimo dell'istituzione) che verrà utilizzato per la generazione dei rispettivi NBN. Ad esempio l'Università di Firenze potrebbe ricevere il sub-namespace UNIFI. Il NBN risultante assumerà la forma: URN:NBN:IT:UNIFI-123456. Le istituzioni che riceveranno il sub-namespace dovranno inoltre necessariamente aderire a Magazzini

Digitali (Vignocchi, Bergamin e Messuti; «Le linee guida per l'accesso aperto alle tesi di dottorato») per il deposito legale dei contenuti da loro prodotti. Questo è evidentemente un processo off-line che deve prevedere la necessaria flessibilità per gestire casi specifici. Ad esempio, nel caso in cui una istituzione abbia digitalizzato dei contenuti, ma non sia in possesso di un IR digitale, le risorse potrebbero essere inviate alla BNCF tramite canale offline (es. CD, DVD, etc.).
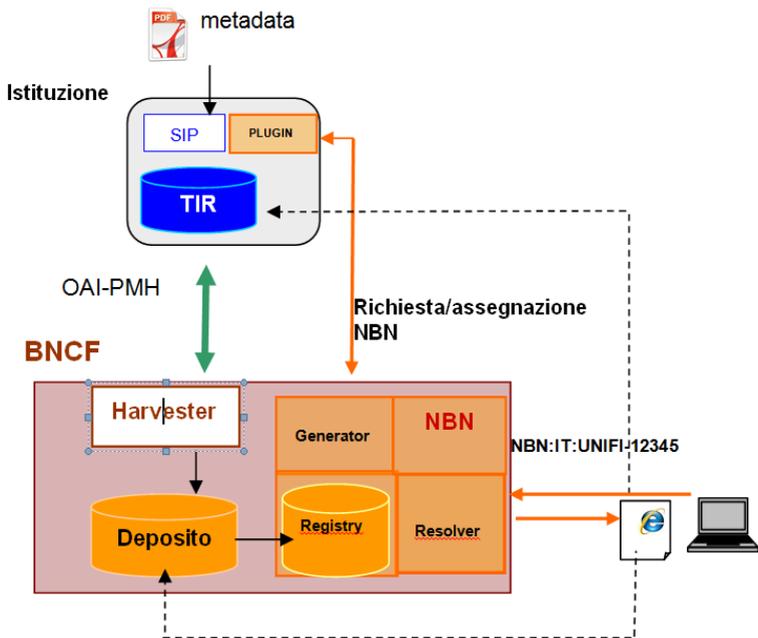
Figura 1: Schema descrittivo di NBN in relazione con il sistema di Deposito Legale

Come si evince dalla figura 1, il sistema NBN è formato da 3 macro blocchi funzionali lato centrale e un plugin lato CP. I tre blocchi

funzionali del sistema centrale sono: il Generatore, il Registro Centrale (RC) e il Resolver. Nei paragrafi seguenti vengono descritti gli elementi dal punto di vista delle loro funzioni principali tralasciando le modalità tecnologiche di implementazione.

Il Registro Centrale NBN

Il Registro Centrale (RC) è il cuore del sistema, il luogo dove vengono mantenute le associazioni NBN-URL. Una delle principali assunzioni di base per il RC è quella di non prevedere la gestione dei metadati descrittivi. Questa decisione si basa sia sulla complessità che sul costo di mantenere aggiornati i metadati presenti nei vari archivi collegati al deposito legale. In questo modo si evita di ricreare un nuovo catalogo e ci si focalizza sul servizio di identificazione, mantenendo invece un puntatore ai metadati autorevoli presenti negli archivi delle istituzioni. La distinzione tra fare un catalogo e un registro è sostanziale. Nel caso del RC di NBN si è provveduto a definire una serie di informazioni gestionali del nome stesso per monitorarne lo stato, la tracciabilità delle modifiche, etc. Il registro quindi mantiene tutte le richieste, le azioni, le modifiche che avvengono sul registro stesso al fine di garantire il trust complessiva del servizio.

Il Generatore NBN

Un NBN viene assegnato a quello che nella terminologia PREMIS[13] viene identificato come *representation*. In altre parole un record formato da più file riceve un solo NBN. Inizialmente il NBN sarà assegnato alle tesi di dottorato visto che questa tipologia di oggetti

[13] Preservation Metadata: Implementation Strategies (PREMIS) è lo standard di metadati per la conservazione promosso da OCLC e RLG, http://www.loc.gov/standards/premis.

è già oggetto di harvesting e deposito da parte di Magazzini Digitali, ma è pensato per essere assegnato a tutte le risorse destinate al deposito legale e alle risorse frutto dei progetti di digitalizzazione condotti dalle biblioteche e da altre istituzioni. La stringa alfanumerica è opaca e non darà alcuna informazione semantica (es. uso di parole specifiche), inoltre, avrà una lunghezza consona alle esigenze di citazione. Si è voluto evitare la complessità d'uso, ad esempio, dell'Universally Unique IDentifiers (UUID) che è problematico a causa della sua lunghezza di 36 caratteri complessivi[14]. Pertanto si è ritenuto opportuno generare un progressivo generale che non sia legato al sub-namespace, ma si basi su un contatore unico.

Il Resolver NBN

Il Resolver[15] è lo strumento che consente l'accesso alla risorsa ma allo stesso tempo anche la sua verifica, che se positiva, consolida e aumenta il trust del sistema, come precedentemente spiegato. Il Resolver è quindi l'interfaccia del servizio verso l'utente e il tipo di risposta che dà è effettivamente il livello di servizio offerto dal sistema. In generale una risoluzione di un NBN prevede la restituzione di una pagina intermedia con

1. i metadati e il link alla risorsa presente nel IR del content provider ritenuto Trusted,
2. le informazioni di stato del NBN e
3. il riferimento alla copia della risorsa presente nel deposito legale.

Questo è il livello di servizio principale dal quale poi è possibile evolvere verso sistemi più sofisticati in accordo alle esigenze della

[14]UUID, http://en.wikipedia.org/wiki/Universally_unique_identifier.
[15]Il dominio di risoluzione sarà http://nbn.depositolegale.it.

comunità. È importante notare che il servizio di deposito legale ha tempi di harvesting diversi rispetto al flusso di pubblicazione delle risorse. L'intero workflow include quindi la presenza di un periodo di latenza tra la richiesta di registrazione del NBN per la risorsa e la sua effettiva archiviazione nel sistema MD. Questa latenza, però non andrà a inficiare il livello di servizio di risoluzione offerto. Infatti si assume che il servizio di risoluzione centrale sia in grado di risolvere il NBN dal momento in cui viene generato, a fronte della richiesta da parte dei CP e non solo dopo che è stato archiviato.[16]

Il plugin per gli IR

Uno degli aspetti più critici di un sistema di identificazione non è la generazione del nome o il processo al termine del quale l'identificativo viene assegnato, ma la sua concreta consegna al richiedente. La consegna implica una sorta di "uscita" dell'identificativo dal sistema generante per andare a inserirsi in uno spazio codificato del richiedente (il posto della targa, la carta di identità, la tesserina del codice fiscale, etc.). A questo proposito l'obiettivo più importante era la consegna contestuale alla richiesta (quindi immediata) del NBN generato, così da entrare subito a far parte dei metadati descrittivi della risorsa. Questo approccio si differenzia, migliorandolo, da quello fino a oggi adottato dalla Biblioteca nazionale tedesca che prevede una attesa di 24/48 ore. Per ottenere questo risultato importante sia in termini di usabilità che di efficienza del sistema è stato sviluppato un apposito plugin, al momento adattabile per il software EPrints[17] che consentirà il dialogo con il sistema centrale tramite web service.

[16] L'interfaccia progettata per essere multilingue presenterà le informazioni necessarie sia sul servizio sia sulle modalità di partecipazione.

[17] È previsto a breve anche lo sviluppo di un plugin per DSpace, che con EPrints costituisce il software più utilizzato dalle università italiane per l'implementazione di open archives.

Il plugin modifica l'interfaccia del sistema di archiviazione usato dall'IR inserendo un elemento attraverso il quale l'operatore sarà in grado di richiedere la generazione del NBN da remoto. Questo strumento, come si vede dalla figura 1 a pagina 15, si inserisce nel workflow di pubblicazione e consente all'operatore di effettuare una richiesta di generazione del NBN prima della validazione finale della risorsa (pubblicazione) e solo per quelle risorse che già si intende inviare al deposito legale.

Sperimentazione e nuove prospettive

Grazie allo sviluppo dei plugin sarà possibile avviare una prima sperimentazione del sistema NBN in collaborazione con il gruppo di lavoro Open Access della Commissione biblioteche della CRUI, coordinato da Roberto Delle Donne. Il gruppo di lavoro ha collaborato con il MiBAC e in particolare con le Biblioteche Nazionali Centrali alla implementazione di una procedura automatizzata per il deposito legale delle tesi di dottorato di ricerca[18] - i primi contenuti digitali a popolare Magazzini Digitali (Vignocchi, Bergamin e Messuti; «Le linee guida per l'accesso aperto alle tesi di dottorato»). Saranno proprio le tesi di dottorato di ricerca di alcune università che hanno già avviato con successo il deposito legale in Magazzini Digitali a ricevere i primi NBN e a testare le funzionalità del sistema.[19] Lo sviluppo di NBN beneficerà anche di un "ecosistema" costituito da una serie di attività correlate con i PI a cui che i partner dell'iniziativa partecipano. Ad esempio, la Fondazione Rinascimento Digitale è leader del workpackage sulla citabilità e gli identificatori nel progetto europeo APARSEN. Tra i risultati del progetto vi è il

[18] http://www.crui.it/homepage.aspx?ref=1891

[19] La CRUI ha incaricato di questa prima sperimentazione l'Alma Mater Studiorum, Università di Bologna e l'Università di Trieste rispettivamente per testare la compatibilità di archivi EPrints e DSpace.

rilascio di un nuovo framework di interoperabilità tra i vari sistemi di identificazione non solo per gli oggetti (NBN, DOI, Handle, etc.) ma anche per gli autori e le istituzioni. L'intento del documento è quello di definire una piattaforma comune di interoperabilità in grado di mettere in relazione queste entità e creare i presupposti per lo sviluppo di servizi avanzati a carattere sia metrico che semantico. Ai lavori partecipano organizzazioni come il CERN, l'Università di Trento, le biblioteche nazionali tedesca e inglese, gli archivi inglesi e molti altri partner. Le attività di sviluppo di NBN stanno già tenendo in considerazione questa prospettiva e i nuovi sviluppi saranno intesi a rendere il sistema, primo in Italia, conforme al framework. Infatti il sistema NBN risponde ai criteri di trust identificati all'interno di APARSEN per i sistemi di PI. I principali criteri a cui NBN corrisponde positivamente sono:

- esistenza di una Registration Autority;
- esistenza di un Resolver accessibile su Internet;
- garanzia dell'unicità dell'identificativo dentro il contesto (dominio) di riferimento;
- garanzia della persistenza del PI (sia a livello di funzionalità che di business model);
- esistenza di policy per gli archivi che ricevono il PI per le proprie risorse;
- risoluzione affidabile;
- disaccoppiamento tra la specifica del PI e la URL di risoluzione.

In particolare, nel confronto con gli attuali sistemi, NBN:IT è del tutto paragonabile al livello di trust ottenuto dal sistema della biblioteca nazionale tedesca sui medesimi criteri.

Un'altra importante iniziativa assunta dalla CRUI partner del consorzio NBN, è la responsabilità della gestione del cosiddetto DOI non commerciale. La CRUI, stipulando un accordo con DataCite, si è attivata per la distribuzione del DOI in particolare alle istituzioni scientifiche che vogliono identificare dataset, risultati della ricerca, articoli, etc. La simultanea presenza nel gruppo di NBN dei responsabili diretti di queste iniziative, consentirà straordinarie sinergie e uno sviluppo coerente di questi servizi con notevoli benefici per l'accesso, il riuso e la diffusione della conoscenza.

Conclusioni

La realizzazione di un sistema di identificazione persistente legato alla digital preservation si propone come parte integrante delle infrastrutture nazionali e internazionali per la ricerca e i beni culturali in primis. L'adozione di identificativi persistenti contribuisce alla definizione dei trust necessaria in particolare per quelle risorse sulle quali la società costruisce e stratifica il nuovo sapere. Occorre dire naturalmente che NBN non intende essere una soluzione alternativa agli identificatori in uso nel mondo dei Linked Data. NBN è un identificatore indipendente dal protocollo di risoluzione ma può essere usato all'interno di Cool URI e offrire un importante servizio anche per i Linked Data (ad esempio il NBN URN:NBN:IT:UNIFI-123456 può essere usato come Cool URI nella forma http://nbn.depositolegale.it/URN:NBN:IT:UNIFI-123456).

Ovviamente, come ogni servizio, il suo successo sarà valutato dal livello di integrazione negli attuali workflow di pubblicazione e produzione di contenuti, dall'accettazione da parte delle comunità e dalla sostenibilità economica. Per favorirne il successo, il consorzio si impegna a tenere in considerazione il contesto internazionale, rece-

pendone le principali direzioni e raccomandazioni, pur adattandole al contesto specifico italiano.

Riferimenti bibliografici

Bazzanella, Barbara, et al. «Persistent Identifiers Interoperability Framework». (2012). (Cit. alle pp. 2, 3).

Bellini, Emanuele e Paolo Ceravolo. «EPICA: Easy Persistent Identifier Common Architecture». *Lecture Notes in Computer Science* 6428. (2010): 626–634. (Cit. a p. 6).

Bellini, Emanuele e Maurizio Lunghi. «Persistent Identifiers for cultural heritage». (2007). <http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf>.

Bellini, Emanuele, et al. «Persistent Identifier Distributed System for Digital Libraries, Information Technology». **proceedings** di World Library And Information Congress: 75th Ifla General Conference And Council, Milan. 2009. (Cit. alle pp. 6, 13).

Bellini, Emanuele, et al. «Persistent Identifiers distributed system for Cultural Heritage digital objects». *iPRESS 2008 Proceedings*. 2008. 55–65. (Cit. alle pp. 5, 6, 13).

Bellini, Emanuele, et al. «Persistent identifiers: jNBN, a JEE application for the management of a national NBN infrastructure». *GARR Conference 2009*. Rome: GARR, 2009. (Cit. alle pp. 6, 13).

Bellini, Emanuele, et al. «Semantics-Aware Resolution of Multi-part Persistent Indentifiers». *Emerging Technologies and Information Systems for the Knowledge Society*. A cura di Miltiadis Lytras, et al. Vol. 5288. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008. 413–422. <http://dx.doi.org/10.1007/978-3-540-87781-3_45>. (Cit. alle pp. 5, 6, 13).

Berners-Lee, Tim. «Cool URIs don't change». *W3C Design Issues*. (1998). <http://www.w3.org/Provider/Style/URI.html>. (Cit. a p. 3).

«CENL Task Force on Persistent Identifiers, Report 2007». (2007). <http://pid.ndk.cz/dokumenty/dalsi-cetba/cenl-task-force-on-persistent-identifiers-report-2007>. (Cit. alle pp. 6, 7).

Daigle, Leslie L., et al. «URN Namespace Definition Mechanisms - RFC 2611». (1999). <https://ebook.tools.ietf.org/html/rfc2611>. (Cit. a p. 11).

«Final Report of the Project Team of CEN/ISSS eGovernment Focus Group, CEN». (2008). <http://www.egovpt.org/fg/Report>. (Cit. a p. 8).

Hakala, Juha. «National Bibliography Numbers as Uniform Resource Names – NBN RFC 3188». (2001). <http://tools.ietf.org/html/rfc3188>. (Cit. alle pp. 11, 12).

«ISO 3297:1986 Documentation - International standard serial numbering (ISSN)». (1986). <http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=8534>. (Cit. a p. 11).

«Le linee guida per l'accesso aperto alle tesi di dottorato». *AIDAinformazioni* 26.3-4. (2008): 73–89. <http://hdl.handle.net/10760/12459>. (Cit. alle pp. 6, 15, 19).

Luhmann, Niklas. «Trust: a mechanism for the reduction of social complexity». *Trust and Power*. A cura di Niklas Luhmann. New York: Wiley, 1979. 4–103. (Cit. a p. 8).

Lynch, Clifford, Cecilia Preston e Ron Jr. Daniel. «Using Existing Bibliographic Identifiers as Uniform Resource Names, RFC 2288». (1998). <http://www.ietf.org/rfc/rfc2288.txt>. (Cit. a p. 11).

Masinter, Larry e Karen Sollins. «Functional Requirements for Uniform Resource Names, RFC 1737». (1994). <http://www.ietf.org/rfc/rfc1737.txt>. (Cit. a p. 11).

Moats, Ryan. «URN Syntax, RFC 2141». (1997). <http://tools.ietf.org/html/rfc2141>. (Cit. alle pp. 3, 11).

«NESTOR Catalogue of criteria for assessing the trustworthiness of PI systems». (2009). <http://files.d-nb.de/nestor/materialien/nestor_mat_13_en.pdf>. (Cit. a p. 2).

«NISO/ANSI Z39.56-1997 Serial Item and Contribution Identifier». (1997). (Cit. a p. 11).

«NISO/ANSI/ISO 2108:1992 Information and documentation - International standard book number (ISBN)». (1992). (Cit. a p. 11).

Solimine, Giovanni. *Controllo bibliografico universale*. Milano: Bibliografica, 1998. (Cit. a p. 11).

Sollins, Karen. «Principles of Uniform Resource Name Resolution IETF, RFC 2276». (1998). <http://www.ietf.org/rfc/rfc2276.txt>.

«The National Libraries Resolver Discovery Service (RDS) - CENL Recommendation». (2007). <http://pid.ndk.cz/dokumenty/dalsi-cetba/the-national-libraries-resolver-discovery-service-a-cenl-recommendation>.

Vignocchi, Marialaura, Giovanni Bergamin e Raffaele Messuti. «Tesi di dottorato: stato dell'arte, iniziative in corso, scenari possibili». *Bibliotime* 13.3. (2010). <http://didattica.spbo.unibo.it/bibliotime/num-xiii-3/vignocchi.htm>. (Cit. alle pp. 6, 15, 19).

Wilson, Frank. «Trust and Identity in Interactive Services: Technical and Societal Challenges». *cc:eGov Organisational change for citizen-centric eGovernment – Think Paper* 11. (2011). <http://zealot.mrnet.pt/govnet/images/articles/5019/PaperTrustandidentityininteractive.pdf>. (Cit. a p. 8).

Winer, Dov. «Persistent Identifiers systems in the Public Administration sector». (2008). <http://www.digitalpreservationeurope.eu/publications/briefs/persisten_identifiers_for_public_administration.pdf>. (Cit. a p. 8).

Emanuele Bellini, Fondazione Rinascimento Digitale
bellini@rinascimento-digitale.it
Chiara Cirinnà, Fondazione Rinascimento Digitale
cirinna@rinascimento-digitale.it
Maurizio Lunghi, Fond. Rinascimento Digitale
lunghi@rinascimento-digitale.it
Cinzia Luddi, Fond. Rinascimento Digitale
cinzia.luddi@gmail.com
Maurizio Messina, Biblioteca Nazionale Marciana
messina@marciana.venezia.sbn.it
Giovanni Bergamin, Biblioteca Nazionale Centrale di Firenze
giovanni.bergamin@bncf.firenze.sbn.it
Raffaele Messuti, Biblioteca Nazionale Centrale di Firenze
raffaele@atomotic.com
Giovanna Cordani, Biblioteca Nazionale Centrale di Roma
g.cordani@bnc.roma.sbn.it
Roberto Delle Donne, Conferenza dei Rettori delle Università Italiane
delledon@unina.it
Francesca Rossi, Conferenza dei Rettori della Universite Italiane
rossi@crui.it
Marialaura Vignocchi, Università di Bologna
marialaura.vignocchi@unibo.it
Stefania Arabito, Università degli Studi di Trieste
arabito@units.it
Jordan Piščanc', Università degli Studi di Trieste
piscanc@units.it

Bellini, E., C. Cirinnà, M. Lunghi, et al. "Il progetto National Bibliography Number Italia (NBN:IT).
Un identificatore persistente a supporto del deposito legale nazionale delle risorse

digitali”. *JLIS.it* Vol.3, n.1 (Giugno/June 2012): 4789-1–4789-26. DOI: 10.4403/jlis.it-4789. Web.

ABSTRACT: L’associazione di un identificatore persistente a una risorsa digitale è considerata oggi una best practice che contribuisce alla risoluzione del problema del reperimento affidabile delle risorse in rete. Attualmente si stanno affermando diverse tecnologie e standard come DOI, Handle e Cool URI con differenti peculiarità e livelli di servizio offerti. Tuttavia, all’interno delle comunità scientifiche e culturali e nel nascente semantic web, persiste la necessità di garantire l’autenticità, la provenienza e in particolare la conservazione delle risorse nel tempo. Il National Bibliography Number (NBN) nasce nel contesto del ruolo e delle responsabilità che le biblioteche nazionali hanno nel campo del controllo bibliografico universale ed è orientato verso l’identificazione di risorse che devono essere conservate nel lungo periodo. Il progetto coordinato dalle biblioteche nazionali di Firenze, Roma e Venezia, dalla Fondazione Rinascimento Digitale e dalla CRUI non è alternativo ai sistemi di PI attualmente in uso, ma ne completa l’offerta.
KEYWORDS: Conservazione digitale; Deposito legale; Identificatori persistenti; National Bibliographic Number; Web semantico

---

Submission: 2012-02-13
Accettazione: 2012-03-21
Pubblicazione: 2012-06-01